

60th
ANNIVERSARY
YEAR

1930

1990

BBC RD 1990/16



Research Department Report

DIGITAL SOUND: Subjective tests on low bit-rate codecs

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E.

**Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION**

DIGITAL SOUND: SUBJECTIVE TESTS ON LOW BIT-RATE CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E.

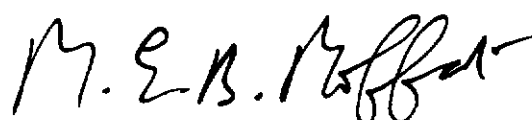
Summary

At the beginning of 1990, BBC Research Department tested four experimental high-quality low bit-rate audio codecs which were under development as part of the Eureka 147 Digital Audio Broadcasting project. The work involved preliminary listening tests, to identify critical test material, followed by formal subjective tests to determine audio quality and error performance.

The listeners could detect some loss of audio quality with all of the codecs, using the most critical material. There were also indications that one of the codecs did not always reproduce the phantom sound sources in their correct positions.

Index terms: *digital audio signals; bit rate reduction; subjective tests*

Issued under the Authority of



Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

Head of Research Department

© British Broadcasting Corporation

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission.

DIGITAL SOUND: SUBJECTIVE TESTS ON LOW BIT-RATE CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E.

1. Introduction	1
2. Brief Details of the Codecs	1
3. Arrangements for the Tests	1
3.1 Arrangements for preliminary tests and the preparation of master recordings	1
3.2 Arrangements for assembling the test sequences	2
3.3 Arrangements for replaying the sequences for the listening tests	2
4. Description of the Tests	2
4.1 Preliminary tests and the preparation of master recordings and test sequences	3
4.2 Tests of basic coding quality	4
4.3 Tests of error performance	5
5. Results of Tests	5
5.1 Results of the basic quality tests	5
5.2 Results of the error tests	7
6. Conclusions	9
7. Acknowledgements	9
8. References	9
Appendix: Details of the Recordings used in the Tests	10

© BBC 2006. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Research & Development except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/> for a copy of this document.

DIGITAL SOUND: SUBJECTIVE TESTS ON LOW BIT-RATE CODECS

N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E.

1. INTRODUCTION

This Report describes subjective tests conducted by BBC Research Department on two transform codecs and two sub-band codecs developed by members of Working Group 2 of the Eureka 147 Digital Audio Broadcasting project*. In this Report the codecs tested are designated 1, 2, 3 and 4.

The purpose of the work was to help the Eureka project by revealing the strengths and weaknesses of the different codecs, and by identifying critical items of programme material for use in listening tests. The tests were of the basic coding quality and of the impairment caused by errors in the bitstream feeding the decoder. The results are of particular interest to the BBC because they indicate the audio quality obtainable at relatively low bit rates at the stage of development reached when the tests were conducted (January 1990). Further technical development has now taken place within Eureka 147, assisted by the results of these tests.

2. BRIEF DETAILS OF THE CODECS

All of the coders and decoders (codecs) exchanged stereo digital audio signals with the signal sources and listening equipment in EBU/AES form. All operated at a sampling frequency of 48 kHz for the tests described in this Report, but some were capable of being used at other sampling frequencies, such as 44.1 kHz and 32 kHz. Codecs 1 and 2 carried the pre-emphasis 'flag' bit from the channel status of the source signal through to the decoder, and reinserted it into the channel status of the EBU/AES signal leaving the decoder.

All codecs operated at 128 kbit/s per mono channel and at a lower rate (96 kbit/s per mono channel for codecs 2, 3 and 4, and 106 kbit/s per mono channel for codec 1). At the time of the tests codec 1 had not been optimised for use at the lower bit rate.

For the purpose of introducing bit errors, the codecs were supplied to the BBC with exclusive OR gates inserted into the connection between coder and decoder. The 'A' and 'B' channels of coder 1 were multiplexed together by interleaving 16 bits at a time from each of the 2 channels. Coder 2 provided no

multiplexing facilities; Two mono links were provided to the decoder, each with an exclusive OR gate. The interface between coder 3 and its decoder operated at a relatively high clock rate (6.144 Mbit/s) and was intended to carry up to 32 channels of audio; for the purpose of these tests only two channels were implemented. The multiplexer in coder 4 used bit-by-bit interleaving to combine the A and B audio signals. All of the error-insertion arrangements provided by the suppliers of the codecs worked with the BBC's truly random event generator. However, because only half the number of error events are applied to the exclusive OR gates connected in codec 2, with errors occurring simultaneously in both left and right channels of the audio, this equipment may have been at some slight advantage in these tests compared with the other codecs. Such an advantage would not be enjoyed in the practical situation where the A and B audio signals would have to be multiplexed together for broadcasting.

All of the codecs provided error protection when operating at the bit rate of 128 kbit/s. The error tests were conducted at this bit rate. Codecs 2 and 4 also included error protection at 96 kbit/s. The data capacity used for this error protection would have meant that fewer bits were provided for coding the audio, with a consequent reduction in the audio quality at the lower bit rate in the absence of errors.

3. ARRANGEMENTS FOR THE TESTS

3.1 Arrangements for preliminary tests and the preparation of master recordings

For preliminary tests, each codec was connected to the signal sources, Compact Disc player or DAT recorder, and to the destinations, a digital-to-analogue converter (DAC) or the DAT recorder (again) via a 4 by 4 EBU/AES switching matrix. Because the Compact Disc player operates at a sampling frequency of 44.1 kHz, a sampling-frequency converter was necessary when signals were sourced from Compact Disc. This was made available for use with all sources and destinations via the matrix.

The interconnection of the various items of digital equipment for preliminary testing and for the making of master recordings is shown in Fig. 1. The EBU/AES switching matrix was a simple arrangement of logic gates, with the appropriate line senders and receivers, with no re-clocking of signals anywhere in the unit. Despite its limitations, no problems occurred

* The tests were conducted with the assistance of M.K.E. Smith and A.J. Stirling

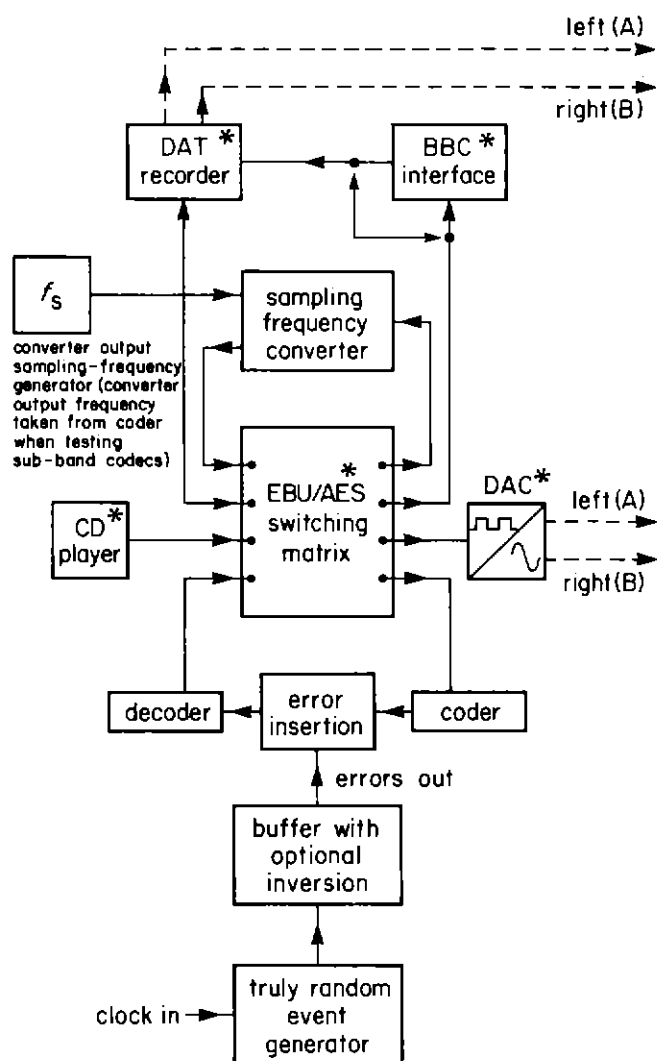


Fig. 1 - Arrangement of digital equipment for making master recordings.

Note: All solid-line connections are EBU/AES
All dotted-line connections are analogue audio

* These items were in the listening room (all others were in the adjacent laboratory)

as a result of its use with the relatively short distances (up to about 10 metres) involved in making the input and output connections. Some problems were encountered, however, with interfacing between individual items of equipment, but these were resolved prior to the preparation of master recordings.

Loudspeaker and headphone listening facilities for the preliminary assessment of quality, and for monitoring during the preparation of master recordings and the compilation of the test sequences were provided using the arrangement of Fig. 2.

3.2 Arrangements for assembling the test sequences

A disc-based digital audio editor¹, developed at BBC Research Department, was used to assemble

material from the master DAT tapes and specially-recorded announcements to form the sequences for the listening tests. The arrangement of the digital equipment for editing is shown in Fig. 3. The analogue signals from the digital-to-analogue converters in the DAT recorder and in the editor were monitored using the arrangement of Fig. 2.

3.3 Arrangements for replaying the sequences for the listening tests

When the test material had been recorded in the final form for the listening tests, the DAT recorder and a high-quality DAC were used in the arrangement of Fig. 4 to replay the recordings to the listeners. Two listening rooms were used; one for listening on BBC monitoring loudspeakers, type LS5/8, and the other for listening on high-quality headphones.

Both listening rooms had a low ambient noise level, representative of good domestic listening conditions. The loudspeaker listening room had a mid-band reverberation time of about 0.3 s and a room volume of about 85 m³. Listening level was adjusted before the start of the tests so that the highest peaks of the music reached 92 dBA at the centre of the front of the listening area.

In the headphone listening room, the same level was set initially using LS5/8 loudspeakers. The headphone level was then set so that subjectively it appeared to be the same as that from the loudspeakers. Finally, the level at the earpiece of one of the headphones was checked using an artificial head². The level at the earpiece was found to be about 1 dB higher than that produced by the loudspeakers, and this setting was retained for the tests[†].

4. DESCRIPTION OF THE TESTS

Preliminary tests took place with each codec prior to the formal subjective tests.

The formal subjective tests fell into two categories; tests of the basic quality of the signal processed by the codecs, and tests of the error performance. Two groups of listeners were involved. One group comprised representatives of the member organisations in the Eureka 147 project; the individuals concerned were from broadcasting organisations, manufacturers and the recording industry. The other group were BBC staff from Research Department and from Radio Broadcasting. The participants were engineers, and most had experience in subjective

[†] It is possible to set headphone levels as much as 10 dB higher than loudspeaker levels using the method described, and the check was made for safety reasons. On this occasion the headphone level was considered satisfactory.

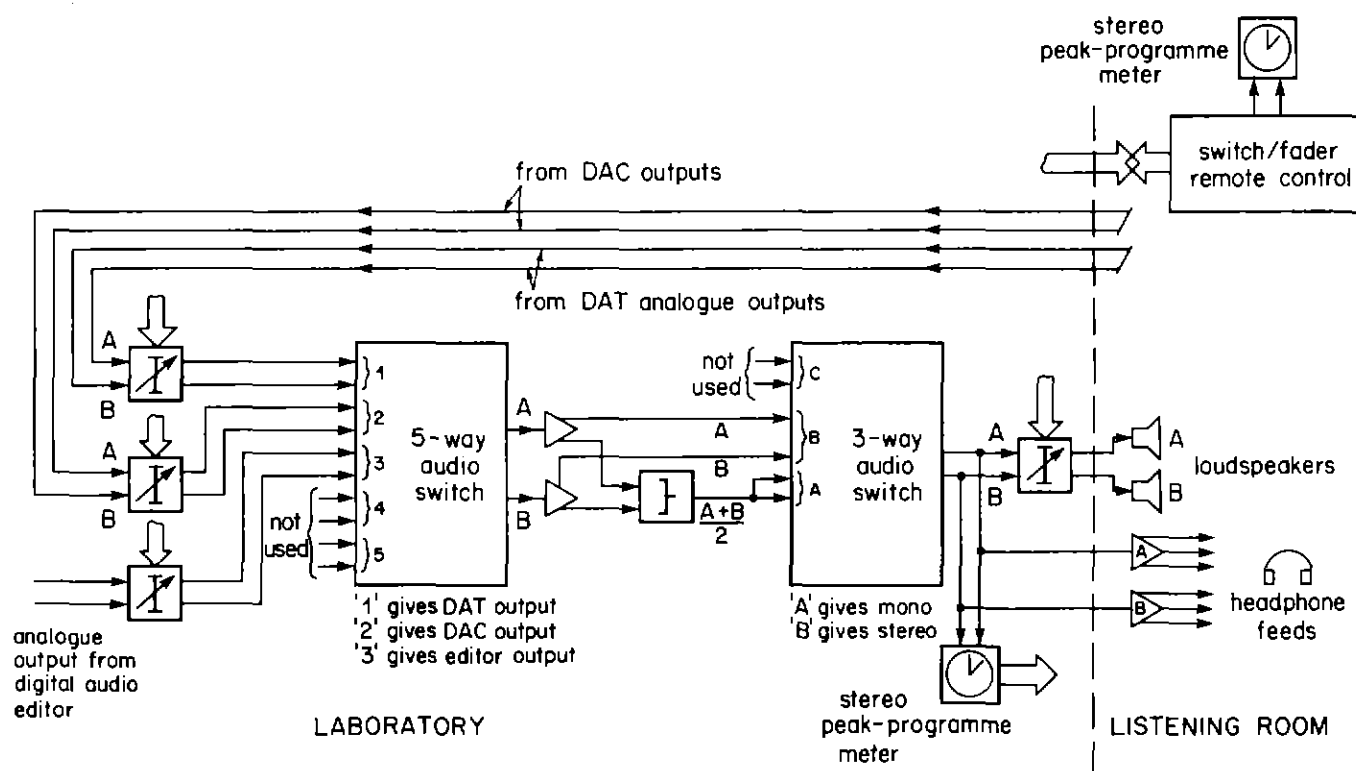


Fig. 2 - Arrangement of analogue equipment during recording and editing.

assessment (but not necessarily in connection with low bit-rate audio coding).

4.1 Preliminary tests and the preparation of master recordings and test sequences

The preliminary tests were conducted principally to determine which programme items were the

most severe test of the basic quality of the various codecs and the range of bit-error ratios over which each codec could realistically be tested. No special study was made of programme material for error testing; solo soprano and pianoforte recordings were used.

Each codec was subjected initially to an

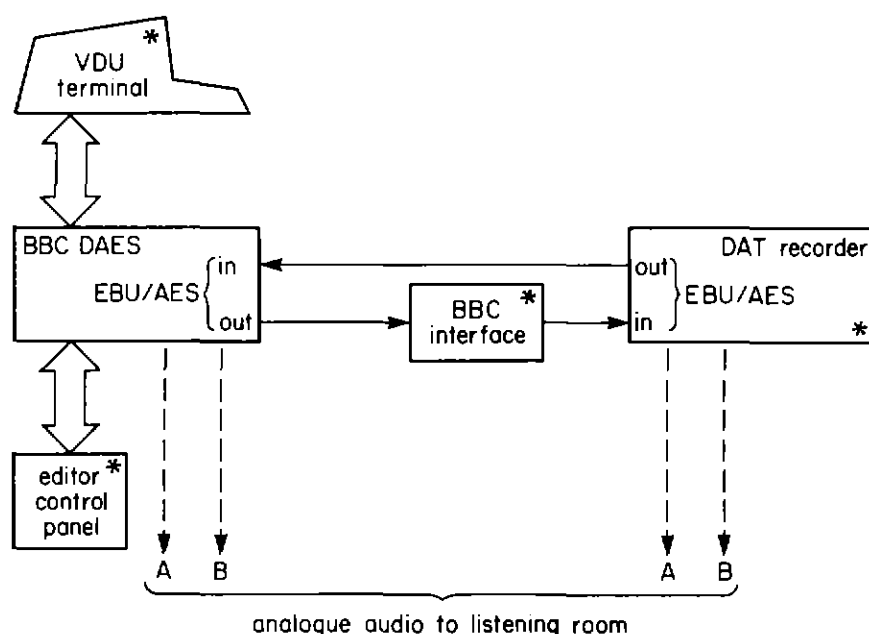


Fig. 3 - Arrangement of digital apparatus for editing test programmes.

Note: All solid-line connections are EBU/AES. All dotted-line connections are analogue audio.

* These items were in the listening room (all others were in the adjacent laboratory)

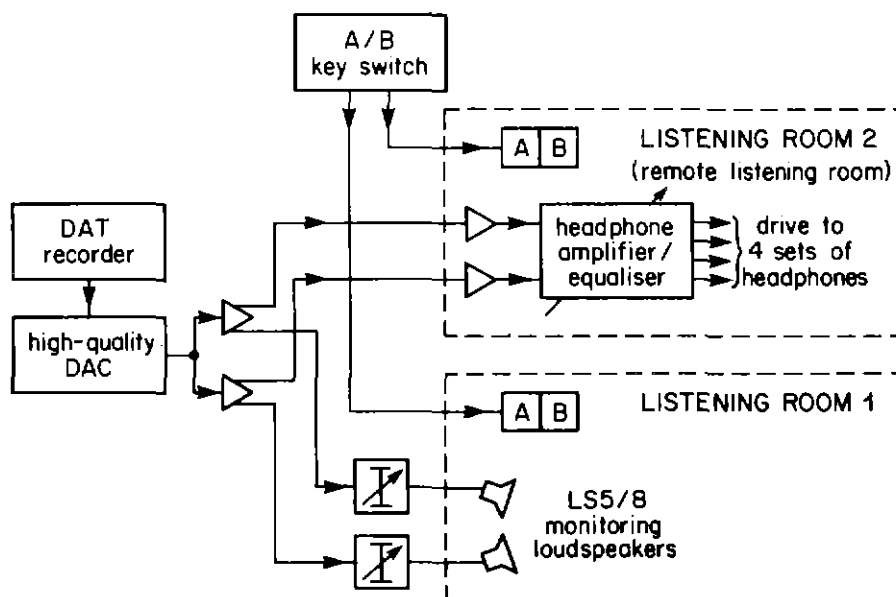


Fig. 4 - Arrangement for replaying recordings to listeners.

informal listening test. Approximate CCIR impairment grades^{3*} were awarded for each programme item, while operating the codec at the lower bit rate (96 or 106 kbit/s, as appropriate). The more critical items were then re-auditioned, and the two most critical items selected for the tests (on the basis of codec performance at 128 kbit/s). By this procedure, up to eight test programme items could have been selected. In fact, codecs 1 and 2 were both found to be sensitive to the solo harpsichord recorded on the EBU SQAM Compact Disc⁴ and codecs 3 and 4 were both sensitive to solo oboe and clarinet *arpeggios* recorded on the same Compact Disc. The test items selected after the preliminary testing were the solo instruments oboe, clarinet and harpsichord from the SQAM Compact Disc, the electronic 'Frère Jacques' signal (also from the SQAM Compact Disc) and a short excerpt comprising an atonal *glissando* played on trumpet and alto saxophone, with accompaniment, from a jazz Compact Disc recording.

Because a number of researchers had found pre-echo effects in low bit-rate audio coding equipment to be shown up by the sound of castanets, a recording of solo castanets, from the SQAM Compact Disc, was included as a test item.

Details of the recordings used for the assessment of both quality and error performance are given in the Appendix.

Test sequences were planned, and master recordings for each codec were made of the material

selected. Recordings of the test sequences were then prepared using the arrangement of Fig. 3. These recordings comprised presentations of signals processed by each codec, interspersed with announcements and preceded by a demonstration of extreme examples of the impairments to be assessed. Three quality test sequences and two error test sequences were prepared, each using two of the test items.

During the preliminary testing, it was found that one of the codecs sometimes distorted the positions of the phantom sound sources created by the reproduced stereo sound signal. The effect was most readily detected using the speech recordings on the SQAM Compact Disc, which are mono. Sometimes the speakers appeared to be situated off-centre, and sometimes just difficult to locate. The codec in question could be reset to cure the problem, but would sometimes re-enter the stereo-distorting mode when the EBU/AES input was switched, or when the decoder muted after a burst of bit errors. Despite attempts during the recording of master tapes to ensure that correct stereo transmission through the codecs was maintained at all times, checks on the recorded test sequences revealed that there were still instances of stereo image problems in the recordings. Prior to the tests, the listeners were warned that such effects could occur, and that whilst they were assessing quality they could, if they wished, indicate that they noticed problems with stereo phantom source location. Any such observations would be noted. They were also assured that they need not feel under any obligation to listen for, or record, such effects if they did not wish to be burdened with this.

* The CCIR 5-point impairment scale is as follows:

- 5 Imperceptible
- 4 Perceptible, but not annoying
- 3 Slightly annoying
- 2 Annoying
- 1 Very annoying

4.2 Tests of basic coding quality

In each test session, lasting between 20 and 25 minutes and comprising 22 tests, listeners assessed the

impairment of the audio signals which had been coded and decoded, and awarded grades using the CCIR 5-point impairment scale³. The recorded material was presented to the listeners four times in each test in the sequence ABAB, where A was always the unprocessed source material and B was normally the signal processed by one of the codecs (but occasionally was the same unprocessed material presented as A). This type of presentation is sometimes termed the 'double stimulus' method. The listeners were not informed of the bit rates in use for each test, nor which tests would comprise the 'testing' of the A signal presented as B (i.e. masquerading as a signal from one of the codecs). At all times during the tests cue lights were used to indicate to the listeners whether it was the A or the B signal which was being reproduced.

The test sequences were planned so that when operating at 128 kbit/s, each codec was tested immediately after a test in which the source signal was presented or after a test in which a codec operating at the lower bit rate (96 or 106 kbit/s) was presented. Furthermore, these particular tests were arranged so that the number following presentations of the source was the same as the number following presentations of codecs at the lower bit rate. The intention in this was to try and minimise any bias of the mean grades for codecs at 128 kbit/s resulting from the degree of impairment of the signal being judged in the previous presentation. It was not possible to provide similar treatment for codecs operating at the lower bit rate.

In each test sequence, two 'dummy' presentations of a codec operating at only 64 kbit/s were included, so that the listeners had to grade a relatively poor signal occasionally. These reference 'bad' signals (sometimes known as the 'lower anchor') always preceded tests in which the source signal was presented as the signal to be graded.

A total of 12 listeners from the Eureka 147 project participated in the formal tests of basic coding quality, providing 12 sets of results for both loudspeaker and headphone tests. One or two additional headphone results were obtained for some tests by BBC and IRT engineers when two of the test sequences were replayed during a check of the listening arrangements before the formal testing commenced. These have been included with the other results. Following the tests with Eureka 147 representatives, a further 22 listeners from the BBC participated in loudspeaker tests.

4.3 Tests of error performance

As already mentioned, errors were provided by a random event generator, which was set to give bit-

error ratios in the range 1 in 10^3 to 1 in 10^2 . Codecs 1 and 2 tended to mute at bit-error ratios in excess of 1 in 10^3 , and were tested only at bit-error ratios up to this figure.

All codecs were error tested only when operating at 128 kbit/s.

The test sessions, each comprising 16 tests, lasted about 20 minutes. As before, the subjective tests used the ABAB sequence of presentation. This time, both A and B signals had been processed by the codec, but the A signal was always error-free. The B signal usually suffered from errors, but occasionally was the same as A. As in the previous tests, the CCIR 5-point impairment scale was used by the listeners, and the listeners were not informed which bit-error ratios were in use for each test, nor which B presentations were the repeated A presentation.

A total of 12 listeners from Eureka 147 participated in the error tests, listening on headphones and on loudspeakers. Tests using loudspeakers continued, using 17 BBC staff after the Eureka participants had completed their tests.

5. RESULTS OF TESTS

At the conclusion of the tests, the individual results were tabulated, and the mean grades awarded by listeners were calculated, for both headphone and loudspeaker listening sessions.

5.1 Results of the basic quality tests

The results of the basic quality tests using headphones are given in Tables 1 and 2; Table 1 contains the results for the codecs when operating at the lower bit rate, and Table 2 contains the results for the codecs at 128 kbit/s. The results for the loudspeaker tests are presented in Tables 3 and 4.

The mean grades obtained in the headphone tests with the codecs operating at the lower bit rates indicate that the solo harpsichord was the signal which suffered the greatest impairment with codecs 1, 2 and 3 (with mean grades of 1.00, 1.89 and 2.43 respectively). The jazz saxophone and trumpet *glissando* was also graded 1.00 when processed by codec 1. Codec 4 gave the lowest mean grade (1.18) with solo oboe. At a bit rate of 128 kbit/s, the castanets proved to be the most critical signal for codecs 1, 2 and 3, with mean grades of 2.75, 2.13 and 3.79 being awarded (codec 3 was graded 3.79 when tested with harpsichord, too). Once again, the oboe represented the most severe test item for codec 4, with a mean grade of 2.79.

Table 1: Results of headphone tests of basic coding quality at low bit rate

Programme material	Codec	Mean grade	Standard deviation	No. of results
Frère Jacques	1	4.14	0.74	13
	2	1.93	0.88	13
	3	4.64	0.61	14
	4	4.57	0.82	14
	Source	4.36	0.72	28
Harpsichord	1	1.00	0	14
	2	1.89	0.71	14
	3	2.43	0.90	14
	4	1.14	0.35	14
	Source	4.45	0.77	28
Clarinet	1	1.92	0.77	12
	2	2.13	0.82	12
	3	2.67	0.92	12
	4	1.08	0.28	12
	Source	4.71	0.45	24
Oboe	1	2.25	0.56	14
	2	3.39	0.66	14
	3	2.96	0.93	14
	4	1.18	0.36	14
	Source	4.86	0.35	28
Jazz sax./tpt.	1	1.00	0	14
	2	2.14	1.12	14
	3	3.07	0.70	14
	4	1.00	0	13
	Source	3.87	1.08	23
Castanets	1	1.54	0.75	12
	2	2.00	1.08	12
	3	3.08	0.86	12
	4	3.83	0.90	12
	Source	4.69	0.61	24

A few results from one of the headphone tests were discarded, because of interference which occurred for a brief period during one of the tests. Fortunately, the problem did not persist, nor was there any recurrence. The loudspeaker tests were not affected.

The mean grades from the loudspeaker tests for codecs operating at the lower bit rate indicate that codec 1 gave the greatest impairment with harpsichord (mean grade 1.15) codec 2 gave the greatest impairment with the electronic 'Frère Jacques' signal (mean grade 2.19) and codecs 3 and 4 gave their greatest impairment with clarinet (mean grades of 2.15 and 1.12, respectively). At 128 kbit/s, codecs 1 and 3 showed the greatest impairment with clarinet (mean grades of 2.79 and 4.07, respectively) codec 2 suffered

Table 2: Results of headphone tests of basic coding quality at 128 kbit/s

Programme material	Codec	Mean grade	Standard deviation	No. of results
Frère Jacques	1	4.69	0.61	13
	2	3.29	1.16	14
	3	4.57	0.49	14
	4	4.36	1.04	14
	Source	4.36	0.72	28
Harpsichord	1	2.89	0.97	14
	2	4.14	0.91	14
	3	3.79	0.94	14
	4	2.89	1.20	14
	Source	4.45	0.77	28
Clarinet	1	3.08	0.95	12
	2	4.21	0.56	12
	3	4.08	0.76	12
	4	4.04	0.83	12
	Source	4.71	0.45	24
Oboe	1	3.39	1.07	14
	2	4.36	0.97	14
	3	4.79	0.41	14
	4	2.79	0.94	14
	Source	4.86	0.35	28
Jazz sax./tpt.	1	4.00	0.85	14
	2	4.21	0.67	14
	3	3.86	0.99	14
	4	3.85	0.86	13
	Source	3.87	1.08	23
Castanets	1	2.75	0.92	12
	2	2.13	0.82	12
	3	3.79	1.07	12
	4	4.33	0.75	12
	Source	4.69	0.61	24

the greatest impairment with the castanets (mean grade 2.99) and codec 4 suffered the greatest impairment (mean grade 3.08) with oboe.

In all of the quality tests, taking the performance of each codec with its most critical signal as the most important criterion, codec 3 consistently suffered the least impairment of any. Whilst other criteria might be applied, it must be emphasised that any averaging of results other than the calculation of the mean grade for each test item is most unlikely to give fair comparison of codec performance. This is because the test material was selected solely to provide two of the most critical test items for each codec, and not a selection of material which was critical for all systems under test. Any averages of results from tests using different programme items are likely to include

Table 3: Results of loudspeaker tests of basic coding quality at low bit rate

Programme material	Codec	Mean grade	Standard deviation	No. of results
Frère Jacques	1	3.91	1.20	34
	2	2.19	1.08	34
	3	4.79	0.47	34
	4	4.63	0.61	34
	Source	4.85	0.39	68
Harpsichord	1	1.15	0.35	34
	2	2.56	1.14	34
	3	2.76	1.11	34
	4	1.38	0.54	34
	Source	4.54	0.67	68
Clarinet	1	1.79	0.69	34
	2	2.21	1.04	34
	3	2.15	0.91	34
	4	1.12	0.32	34
	Source	4.68	0.60	68
Oboe	1	2.32	0.96	34
	2	3.38	0.87	34
	3	2.35	0.76	34
	4	1.18	0.71	34
	Source	4.69	0.67	68
Jazz sax./tpt.	1	1.35	0.68	34
	2	3.06	1.37	34
	3	3.07	0.91	34
	4	1.19	0.45	34
	Source	4.71	0.62	68
Castanets	1	3.01	1.24	34
	2	2.87	1.31	34
	3	4.38	0.59	34
	4	4.40	0.78	34
	Source	4.76	0.49	68

Table 4: Results of loudspeaker tests of basic coding quality at 128 kbit/s

Programme material	Codec	Mean grade	Standard deviation	No. of results
Frère Jacques	1	4.53	0.74	34
	2	3.31	1.03	34
	3	4.74	0.61	34
	4	4.71	0.61	34
	Source	4.85	0.39	68
Harpsichord	1	3.25	1.10	34
	2	4.41	0.65	34
	3	4.17	0.71	34
	4	3.63	1.15	34
	Source	4.54	0.67	68
Clarinet	1	2.79	1.02	34
	2	4.25	0.84	34
	3	4.07	0.73	34
	4	3.32	0.96	34
	Source	4.68	0.60	68
Oboe	1	3.97	0.92	34
	2	4.41	0.81	34
	3	4.65	0.80	34
	4	3.08	0.94	34
	Source	4.69	0.67	68
Jazz sax./tpt.	1	4.00	0.84	34
	2	4.35	0.72	34
	3	4.12	0.87	34
	4	3.85	0.86	34
	Source	4.58	0.70	68
Castanets	1	4.35	0.72	34
	2	2.99	1.25	34
	3	4.50	0.65	34
	4	4.78	0.47	34
	Source	4.76	0.49	68

some items which are non-critical to some (but not all) of the codecs.

In Section 4.1 above, attention has been drawn to the displacement of the phantom sound sources created by the stereo signal when one of the codecs was processing some of the test items. Only in a relatively small number of tests were possible stereo problems indicated by the listeners on their test forms. Most of the indications of possible stereo distortion occurred during the headphone tests.

It would appear that there are good reasons for devising methods for subjectively testing audio systems for this type of impairment. The CCIR has formed a Working Party (WP 11E) responsible for looking into new requirements and methods for subjective testing.

5.2 Results of the error tests

The results of the tests of codecs with errors introduced into the bit stream prior to decoding are presented in Tables 5 and 6. The results for headphone and for loudspeaker tests give mean grades which are consistently closer than those from the quality tests, and in only two tests do the headphone and loudspeaker mean grades differ by more than 0.5 grade (headphones giving the lower grade in both cases).

The more critical loudspeaker results (those for the solo soprano) have been used to plot a number of graphs of mean grade as a function of bit-error ratio. These are shown in Fig. 5. In these tests, codec 3 appears to have achieved the best performance.

Table 5: Results of headphone tests of error performance

Codec 1				
Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	9.8 in 10^4	1.00	0	12
	1.2 in 10^4	1.58	0.64	12
	1.5 in 10^5	2.54	0.92	12
	No errors	4.79	0.38	12
Pianoforte	9.8 in 10^4	1.00	0	12
	1.2 in 10^4	1.46	0.56	12
	1.5 in 10^5	3.13	0.96	12

Codec 2

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	9.8 in 10^4	1.83	0.80	12
	1.2 in 10^4	2.96	0.90	12
	1.5 in 10^5	4.75	0.43	12
Pianoforte	9.8 in 10^4	1.92	1.02	12
	1.2 in 10^4	3.63	0.54	12
	1.5 in 10^5	4.83	0.37	12
	No errors	4.58	0.49	12

Codec 3

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	7.8 in 10^3	1.46	0.63	12
	9.8 in 10^4	2.63	0.79	12
	1.2 in 10^4	3.46	0.80	12
	1.5 in 10^5	4.88	0.30	12
Pianoforte	7.8 in 10^3	1.29	0.52	12
	9.8 in 10^4	3.25	1.18	12
	1.2 in 10^4	4.42	0.61	12
	1.5 in 10^5	4.79	0.38	12
	No errors	4.92	0.28	12

Codec 4

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	7.8 in 10^3	1.17	0.37	12
	9.8 in 10^4	1.67	0.62	12
	1.2 in 10^4	3.13	1.04	12
	1.5 in 10^5	3.88	0.58	12
	No errors	4.83	0.37	12
Pianoforte	7.8 in 10^3	1.04	0.14	12
	9.8 in 10^4	2.17	0.72	12
	1.2 in 10^4	3.79	0.63	12
	1.5 in 10^5	4.63	0.46	12

Table 6: Results of loudspeaker tests of error performance

Codec 1				
Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	9.8 in 10^4	1.00	0	29
	1.2 in 10^4	1.34	0.60	29
	1.5 in 10^5	2.29	0.64	29
	No errors	4.79	0.38	29
Pianoforte	9.8 in 10^4	1.00	0	29
	1.2 in 10^4	1.28	0.43	29
	1.5 in 10^5	3.21	0.98	29

Codec 2

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	9.8 in 10^4	1.48	0.68	29
	1.2 in 10^4	2.59	0.92	29
	1.5 in 10^5	4.81	0.44	29
Pianoforte	9.8 in 10^4	1.57	0.76	29
	1.2 in 10^4	3.40	0.69	29
	1.5 in 10^5	4.83	0.53	29
	No errors	4.84	0.44	29

Codec 3

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	7.8 in 10^3	1.17	0.44	29
	9.8 in 10^4	2.40	0.79	29
	1.2 in 10^4	3.98	0.72	29
	1.5 in 10^5	4.83	0.35	29
Pianoforte	7.8 in 10^3	1.14	0.37	29
	9.8 in 10^4	4.33	0.67	29
	1.2 in 10^4	4.72	0.57	29
	1.5 in 10^5	4.69	0.59	29
	No errors	4.86	0.34	29

Codec 4

Programme material	Bit-error ratio	Mean grade	Standard deviation	No. of results
Soprano solo	7.8 in 10^3	1.00	0	29
	9.8 in 10^4	1.33	0.48	29
	1.2 in 10^4	2.88	0.88	29
	1.5 in 10^5	3.78	0.71	29
	No errors	4.81	0.53	29
Pianoforte	7.8 in 10^3	1.00	0	29
	9.8 in 10^4	2.34	0.76	29
	1.2 in 10^4	4.09	0.73	29
	1.5 in 10^5	4.53	0.76	29

However, it must be emphasised that random errors do not necessarily represent very accurately processes which might be encountered in a real digital audio broadcasting system.

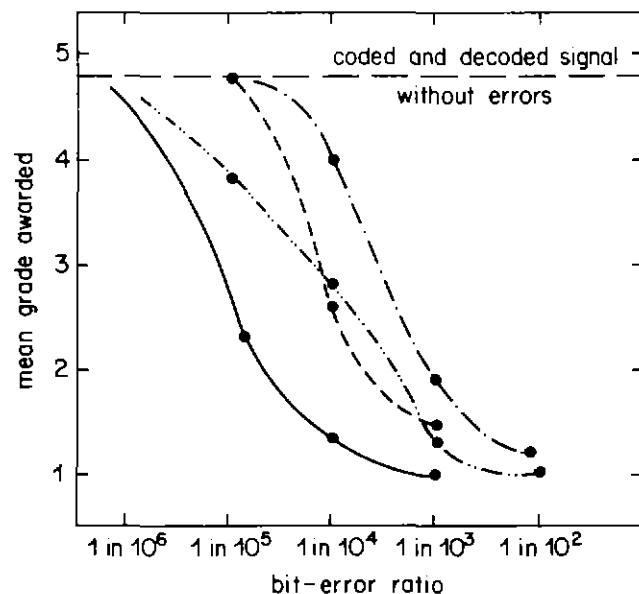
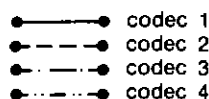


Fig. 5 - Graphs of mean grade as a function of bit-error ratio for the overall loudspeaker results, using the solo soprano test material.



6. CONCLUSIONS

Subjective tests have been conducted on four low bit-rate digital audio codecs developed by members of the Eureka 147 Digital Audio Broadcasting project. None of the codecs was capable of conveying the most critical signals without any perceptible impairment, but there were differences in the degree of subjectively-assessed impairment in tests both of basic coding quality and of resistance to the effects of bit errors. The tests were not intended to provide comparisons between the different codecs, but rather to help the individual organisations involved assess the strengths and weaknesses of their codecs, and identify some of the most critical items of test material. However, whenever results such as these are reported, comparisons are likely to be drawn between the results for different codecs, and it appears that the codec designated 3 introduced the least impairment into the most critical signals and was also the most resistant to the effects of bit errors.

Inevitably, when subjective tests are conducted on the scale of those reported here, one can draw some conclusions concerning the testing methods as

well as the performance of the systems under test. One interesting conclusion to be drawn concerning the testing methods is that there are some types of listening test for which headphones may be better suited than loudspeakers, and *vice versa*. In particular, headphones seem to be better than loudspeakers for tests in which pre-echo is the predominant impairment.

In a few tests, some listeners observed a displacement of the phantom sound sources from their correct positions. Signals from centrally-placed single sound sources, particularly the speech items on the SQAM Compact Disc, revealed some problems with phantom source location. However, further study is necessary to find the most suitable material for revealing stereo problems. As yet, there is no universally-accepted method for subjectively testing stereo distortion of this kind, but it must be a matter of some urgency to define testing methods for assessing distortion of the stereo effect.

7. ACKNOWLEDGEMENTS

The author would like to thank the Centre Commun d'Etudes de Télédiffusion et Télécommunications (CCETT), Deutsche Thomson-Brandt GmbH (DTB), the Fraunhofer Gesellschaft (FhG), the Institut für Rundfunktechnik (IRT), Philips and Decca, for the loan of equipment and for the participation of their staff in listening tests. He also wishes to acknowledge the assistance given by Messrs. H. Schott (FhG), R. Seeman (DTB), J.Y. Roudaut (CCETT), J.Y. Leseure (CCETT), R. Sedlmeyer (IRT), M. Link (IRT) and A. Brefort (IRT) during the preparations for the tests.

8. REFERENCES

1. KIRBY, D.G., 1988. The realisation and exploitation of a random-access digital audio editor for broadcast use. AES 85th Convention, November 1988 (Los Angeles). AES Preprint No. 2710.
2. MATHERS, C.D. and LANSDOWNE, K.F.L., 1979. Hearing risk to wearers of circumaural headphones: An investigation. BBC Research Department Report No. BBC RD 1979/3.
3. CCIR, 1986. Recommendation 562. Subjective assessment of sound quality. Sixteenth Plenary Assembly, Dubrovnik, 1986. Green Book, Vol. X - Part 1, pp 275-279.
4. EBU, 1988. SQAM-sound quality assessment material: recordings for subjective tests. EBU Document Tech. 3253 (including the SQAM Compact Disc).

APPENDIX

Details of the Recordings Used in the Tests

EBU SQAM Compact Disc (*track and index numbers in brackets*)

- | | |
|------------------------------------|--------|
| 1. Frère Jacques (electronic tune) | (7.1) |
| 2. Harpsichord <i>arpeggio</i> | (40.1) |
| 3. Clarinet <i>arpeggio</i> | (16.1) |
| 4. Oboe <i>arpeggio</i> | (14.1) |
| 5. Castanets | (27.1) |
| 6. Soprano | (44.1) |
| 7. Pianoforte (first 14 seconds) | (60.1) |

Compact Disc: 'In All Languages' by Ornette Coleman (**'the Original Quartet and Prime Time'**) **Dreams 008 (1987).**

1. Alto Saxophone, trumpet and accompaniment (bass, percussion) on track 7 of the disc, starting at timecode 0 sec. and ending at timecode 22.0 sec.